

## G.8. Los datos bibliométricos extraídos de registros de PubMed no son fiables

Por Carlos B. Amat y Alfredo Yegros-Yegros

23 diciembre 2010

**Amat, Carlos B.; Yegros-Yegros, Alfredo.** "Los datos bibliométricos extraídos de registros de PubMed no son fiables". *Anuario ThinkEPI*, 2011, v. 5, pp. 223-229.



**Resumen:** La realización de estudios bibliométricos para caracterizar la investigación sobre un tema exige una selección muy cuidadosa de los registros de base de datos que representan los trabajos. Cuando los estudios se realizan en un contexto nacional o comparan la producción científica de varios países, es imprescindible que las fuentes utilizadas expresen de forma completa la filiación institucional de los autores. El uso de Medline (PubMed) para la selección de trabajos es rutinario, pero provoca grandes inexactitudes porque sólo incluye los datos de filiación del primer firmante. Es necesario su uso en combinación con otras fuentes. Scopus, que comparte ciertos elementos con Medline, como

el PMID y los descriptores (Indexterms) ofrece resultados fiables y permite la necesaria combinación de criterios para la selección de documentos y su análisis bibliométrico.

**Palabras clave:** Medline, Bibliometría, Bases de datos bibliográficas, PubMed.

**Title:** *Data extracted from PubMed records are not reliable*

**Abstract:** Bibliometrics on a research subject needs a careful selection of bibliographic records to be retrieved and downloaded. Whenever national production is evaluated or international output is compared, it is essential to have full information on institutional and country affiliation of every author. As Medline offers only the affiliation of the first author, it is mandatory to combine its results with other databases. Scopus is the obvious choice because it offers common elements in its data schema (such as PMID and Indexterms) and complete information on authors' affiliation.

**Keywords:** Medline, Bibliometrics, Bibliographic databases, PubMed.

### Introducción

**LA BIBLIOMETRÍA se ha convertido en las últimas décadas en una herramienta habitual para el análisis de la actividad científica y el desarrollo técnico.**

Además de una bibliometría básica eminentemente metodológica, y de la de evaluación, que se aplica a la gestión y la política científica, la bibliometría temática rinde excelentes frutos en la caracterización de líneas y frentes de investigación y de trayectorias tecnológicas.

El esquema básico de un análisis bibliométrico pasa por la selección de la base de datos a emplear como fuente, la recuperación de los registros que cumplen determinados requisitos y el análisis cuantitativo de las variables que se desea estudiar. La selección de la base de datos está basada en el tema a investigar, el ámbito geográfico y la

forma en que cada base de datos estructura los registros bibliográficos que contiene.

El hecho de que una base de datos constituya referencia obligada y fuente de información imprescindible en un campo de actividad científica no la convierte en una fuente bibliométrica fiable. Esta nota tiene por objetivo presentar algunos ejemplos de dicha falta de fiabilidad. Se propone buscar un consenso sobre una metodología sólida para el análisis de la actividad científica a través de sus publicaciones.

---

**"El recurso a Medline para la selección de trabajos sometidos a análisis bibliométricos hace tiempo que se convirtió en rutina"**

---

## Un ejemplo de peso

Recientemente se ha publicado el penúltimo de una serie de trabajos bibliométricos sobre algunos aspectos de la investigación médica española. Se ha elegido la base de datos *Medline* como fuente para el estudio de la producción científica española sobre la obesidad. Como único requisito, los registros debían contener el descriptor "Obesity". Al extender la búsqueda al período 1988-2007, se recuperan 58.325 documentos y, de los 50.120 con datos de filiación institucional, se analizan los 1.407 trabajos identificados como españoles (Vioque et al., 2010).

Cuando se recuperan conjuntos de trabajos para su posterior análisis bibliométrico se emplean perfiles guiados por tres criterios: el tema, las características de sus autores (que incluyen desde su perfil disciplinario a su adscripción institucional y geográfica) o el medio en que se han publicado, además de otros como plazos temporales. Los autores mencionados han empleado sólo un perfil temático, y eso no es suficiente.

Por ejemplo, no han recuperado el original *In vivo effects of CGP-12177 on the expression of leptin and uncoupling protein genes in mouse brown and white adipose tissues* (PMID 10805498), porque no ha sido indizado por los términos mencionados aunque apareció en el *International journal of obesity*. Sin embargo, es relevante y se debe contabilizar, al igual que los demás trabajos de grupos españoles publicados en las revistas especializadas en la investigación de la obesidad, pero no por pertenecer a un "medio" especializado.

En cuanto al perfil de los grupos de autores, muchos de los trabajos del CIBER-OBN pueden parecer ajenos a la temática que mencionamos. <http://www.ciberobn.es/webciber>

Sin embargo, son trabajos de los 27 grupos que integran el Centro de Investigación Biomédica

en Red de Fisiopatología de la Obesidad y Nutrición. Colectivamente, estos grupos españoles publicaron 351 trabajos de investigación en 2007. Vioque y colaboradores (p. 227) contabilizan ese mismo año sólo 213 publicaciones para toda España. El problema no es la formulación de las estrategias de recuperación sino la elección de la fuente, en este caso *Medline*, para la selección de los registros.

## Las limitaciones de Medline

Los trabajos de investigación son obra de diversos grupos, pero *Medline* incluye en la mayoría de sus registros sólo la afiliación del primer firmante y en determinados períodos ni siquiera ha incorporado a sus registros todos los autores de cada trabajo (*Bibliographic Services Division*, 2010).

En 1990 se publicaron 4.856 trabajos de investigación biomédica con participación de grupos españoles. El dato, de elaboración propia, procede de la *Web of Science (WoS)* y se limita a los artículos originales, las revisiones y las cartas. La cifra de *Medline* para ese mismo año es más modesta: 2.915 publicaciones. En aquellos 4.856 originales, revisiones y cartas, participaba una media de 0,19 instituciones extranjeras. En 2007, el último año del período de estudio de Vioque y colaboradores (2010), el promedio se ha elevado a 1,12 instituciones extranjeras participantes en 16.120 trabajos españoles de investigación biomédica.

*Medline* incluye ese año 15.728 trabajos españoles, es decir, trabajos cuyo primer firmante está afiliado a un centro español. Pero los colaboradores españoles no siempre son los primeros en el orden de firma (Jiménez-Contreras et al., 2010). Para averiguar en qué orden se mencionan las instituciones españolas en los trabajos contamos con otro ejemplo.

*Ciberer*, otro de los centros de investigación en red, está constituido por 61 grupos españoles de investigación de enfermedades raras. Venimos realizando un seguimiento bibliométrico de la producción científica de esos grupos con datos a partir de 2000 y algunos de los resultados obtenidos son relevantes para esta nota.

En el período 2007-2009 (la fuente también

The screenshot shows the PubMed.gov search interface. The search bar contains the query: "Antioesity[Title] AND properties[Title] AND two[Title] AND African[Title] AND". Below the search bar, it indicates "We found 1 article by title matching your search:". The article details are as follows:

- Phytother Res.** 2007 Dec;21(12):1253-5.
- Antioesity properties of two African plants (Afromomum meleguetta and Spilanthes acmella) by pancreatic lipase inhibition.**
- Ekanem AP, Wang M, Simon JE, Moreno DA.**
- Institute of Oceanography, University of Calabar, P.M.B. 1115, Calabar, Nigeria. albertekanem@yahoo.com**
- Abstract**
- Ethanol extracts of seeds of *Afromomum meleguetta* and flower buds of *Spilanthes acmella* presented pancreatic lipase inhibitory activities in a concentration related manner under in vitro conditions. The two plants were extracted with 70% ethanol by sonication, fractionated on silica gel and tested at concentrations in the range 0.75-2.0 mg/mL. Lipase inhibitory activities of 90% and 40% were observed in *A. meleguetta* and *S. acmella*, respectively. The two plants have potentials as candidates for weight reduction and obesity control.
- Copyright (c) 2007 John Wiley & Sons, Ltd.
- PMID: 17705140 [PubMed - indexed for MEDLINE]
- Publication Types, MeSH Terms, Substances
- LinkOut - more resources

Búsqueda en PubMed con identificación de la adscripción institucional para el primer autor

es WoS) los grupos integrados en *Ciberer* han publicado un total de 1.570 trabajos; 1.175 de ellos son originales, revisiones o cartas. El número de afiliaciones institucionales es de 4.279 y la cuarta parte (1.208) corresponden a instituciones españolas. Pues bien, sólo en el 16,68% de los trabajos la institución del primer firmante es española. El 83% restante cuenta con participación de las instituciones españolas integradas en *Ciberer*, pero no se podrían localizar puesto que estas instituciones no figuran en el campo correspondiente de *Medline*. En el marco de estudios sobre la producción científica en biomedicina de alguna región española, la pérdida ha llegado a

calcularse en una tercera parte de los trabajos (**Costas; Moreno; Bordons**, 2008).

El recurso a *Medline* para la selección de trabajos luego sometidos a análisis bibliométricos hace tiempo que se convirtió en rutina. La tabla 1 indica el año de publicación, el tema, la cobertura temporal y el identificador de algunos análisis bibliométricos de la investigación española que toman *Medline* como fuente de selección de los registros.

No son los únicos: grupos tan solventes como el de **Sanz-Casado** han seguido sorprendentemente la misma práctica (**Lascurain-Sánchez et al.**, 2008). Pero el uso de las bases de datos que sólo

Año	Tema	PMID	Revista
1995	Desigualdades en salud (1980-1994)	8582797 <sup>b</sup>	<i>Gac Sanit</i>
1996	Evaluación de calidad sanitaria (19983-1992)	20524272 <sup>b</sup>	<i>Gac Sanit</i>
1999	Atención primaria (1990-1997)	10408133	<i>Aten Primaria</i>
1999	Anestesiología (1989-1998)	10613074	<i>Rev Esp Anesthesiol Reanim</i>
1999	Tabaquismo (1970-1996)	10098992	<i>Eur J Epidemiol</i>
2000	Autores y revistas de gran impacto (1993-1997)	10846685	<i>Med Clin (Barc)</i>
2000	Sistema respiratorio (1987-1998)	11093882	<i>Med Clin (Barc)</i>
2000	Vacunas infantiles (1990-1998)	11268542 <sup>c</sup>	<i>Aten Primaria</i>
2001	Comunicación y salud (1995-2000)	11334591 <sup>b</sup>	<i>Aten Primaria</i>
2001	Psicología (1989-1998)	11723637	<i>Span J Psychol</i>
2001	SIDA (1991-1999)	11734169 <sup>a</sup>	<i>Med Clin (Barc)</i>
2002	Pediatría (1996-2001)	12139871	<i>An Esp Pediatr</i>
2003	Otorrinolaringología (1998-2002)	14992119	<i>Acta Otorrinolaringol Esp</i>
2003	Anestesiología y resucitación (1983-1995)	14753139 <sup>b,c</sup>	<i>Rev Esp Anesthesiol Reanim</i>
2004	Enfermedades infecciosas (1991-2001)	14757004 <sup>a</sup>	<i>Enferm Infecc Microbiol Clin</i>
2004	Diagnóstico por la imagen (1994-1998)	15373986 <sup>b</sup>	<i>Rev Esp Cardiol</i>
2004	Virus de la hepatitis C (1980-2002)	15461937	<i>Gastroenterol Hepatol</i>
2005	Sistema respiratorio (1990-2002)	15898882	<i>An Med Interna</i>
2005	Microbiología (1990-2002)	16159540 <sup>a</sup>	<i>Enferm Infecc Microbiol Clin</i>
2005	Diagnóstico ecográfico en obstetricia (1994-1998)	sin número <sup>b</sup>	<i>Radiología</i>
2005	Agentes antimicrobianos (1990-2002)	15915233 <sup>a</sup>	<i>Rev Esp Quimioter</i>
2005	Enfermedades raras (1990-2000)	16185632	<i>Med Clin (Barc)</i>
2006	Dermatología (1998-2000)	sin número <sup>a</sup>	<i>Actas Dermosifiliogr</i>
2006	<i>Helicobacter pylori</i> (1988-2003)	16792455	<i>Rev Esp Enferm Dig</i>
2007	Bioética (1966-2003)	17916307 <sup>a</sup>	<i>Gac Sanit</i>
2008	Dermatología (1997-2006)	18501169	<i>Actas Dermosifiliogr</i>
2008	Tuberculosis en el mundo (1997-2006)	19017458 <sup>a</sup>	<i>Int J Tuberc Lung Dis</i>
2008	Resonancia magnética (2001-2007)	18424037	<i>Eur J Radiol</i>
2009	Tuberculosis (1997-2006)	19467751 <sup>a</sup>	<i>Arch Bronconeumol</i>
2010	Psicofarmacología (1987-2007)	20480716	<i>Span J Psychol</i>
2010	Obesidad (1988-2007)	20381211 <sup>a</sup>	<i>Gac Sanit</i>
2010	Dolor de cuello y espalda (1992-2006)	20044103 <sup>c</sup>	<i>Med Clin (Barc)</i>

Tabla 1. Algunos análisis bibliométricos de la investigación española que toman *Medline* como fuente de selección de los registros

a. trabajos de un mismo grupo - b. trabajos que combinan *Medline* e *Índice Médico Español* - c. trabajos que combinan *Medline* y *Embase*

proporcionan la dirección para correspondencia o una afiliación está explícitamente desaconsejado en bibliometría (**Glänzel, 1996**).

*Medline* no incluye las filiaciones institucionales de todos los autores de cada artículo, no lista las referencias bibliográficas de cada trabajo ni permite relacionar unos con otros a través de su relación de cita directa, y no informa de su popularidad, según el número de citas que reciben. En otras palabras: *Medline* no es una base de datos bibliométrica, por mucho que se haya defendido su utilidad (**Pestaña, 1997**).

El resultado invariable de su empleo en los trabajos de la tabla 1 y en otros es una representación disminuida de la participación o contribución de grupos españoles en línea o en temas de investigación biomédica. La distribución de las publicaciones por países (**Vioque et al., 2010a; Vioque et al., 2010b**) o el estudio de la colaboración institucional vendrán lastrados por inexactitudes inadmisibles.

“El uso de las bases de datos que sólo proporcionan la dirección para correspondencia o una afiliación está explícitamente desaconsejado en bibliometría”

La combinación de fuentes como solución

Register for free at <https://www.scipedia.com> to download the version without the watermark

*Scopus* (Elsevier) es una base de datos multidisciplinar de resúmenes y citas aparecida en 2004, aunque su cobertura se extiende retrospectivamente e incluye más de 16.000 revistas científicas. Al menos 12 universidades españolas, además de los más de 100 centros del CSIC, han suscrito el acceso a la misma. Su combinación con *Medline* se viene empleando para delimitar conceptualmente un campo (*Medline*) y para obtener información exhaustiva sobre las referencias contenidas en los trabajos seleccionados (*Scopus*) en lo que se ha empezado a denominar método multibase de datos (**Strotmann; Zhao, 2010**).

Pero los hechos más interesantes y ajustados a esta nota son los siguientes:

- *Scopus* incorpora todos los registros de *Medline*;
- incluye, en su elemento “Indexterms”, los descriptores asignados a los registros de *Medline*;
- cada registro dispone del correspondiente identificador (PMID)
- los registros contienen la afiliación de todos los autores de cada trabajo.

Con *Scopus* es posible, por tanto:

- 1) seleccionar trabajos que traten un tema y se hayan realizado en instituciones de un país;
- 2) combinar los registros recuperados en *Scopus* con los correspondientes en *Medline* a través del PMID (también del DOI para registros recientes);
- 3) llevar a cabo diferentes análisis.

Siguiendo la sintaxis de *Scopus*, la consulta avanzada “INDEXTERMS (obesity) AND (PUBYEAR BEF 2008) AND (PUBYEAR AFT 1987) AND AFFILCOUNTRY (Spain)” arroja como resultados 2.713 trabajos (7 de Julio de 2010). De ellos, el titulado *Antiobesity properties of two African plants (Afromomum meleguetta and Spilanthes acmella) by pancreatic lipase inhibition* (PMID 17705140) se ha realizado en el Instituto de oceanografía de la Universidad de Calabar (sudeste de Nigeria) según *Medline*.

En realidad, el estudio ha contado con la participación del Programa de productos vegetales naturales de la Universidad de New Jersey (EUA) y del Departamento de ciencia y tecnología de alimentos de la Universidad de Murcia (España). Un hermoso ejemplo de trabajo en colaboración que *Medline*, sin embargo, pasa por exclusivamente africano.

La selección de publicaciones científicas para su análisis bibliométrico se debe realizar de alguno de estos modos:

- 1) empleo de una base de datos bibliométrica;
- 2) combinación de dos o más bases de datos bibliométricas;
- 3) construcción *ad hoc* de bases de datos que hayan extraído de otras los datos bibliográficos y bibliométricos y deriven los oportunos indicadores o variables a partir de su procesamiento.

“El hecho de que una base de datos constituya referencia obligada en un campo no la convierte en una fuente bibliométrica fiable”

*Web of Science* (WoS) y *Scopus* son las preferidas para la primera opción (**Falagas et al., 2008**). De hecho, el propio grupo de **Vioque** eligió WoS para su análisis geográfico de la investigación del sida (**Ramos et al., 2005**). La segunda opción ha rendido buenos frutos para el grupo de **Rosa Casado**, que emplea hasta 10 bases de datos en la selección de los trabajos que analiza (**Casado-Mejía; Ruiz-Arias; Solano-Parés, 2009**) y para el grupo de **Manuel Posada**, que combina con otras cuatro bases los elementos bibliométricos



añadidos recientemente a *PsycINFO* (**Belinchón-Carmona**, 2010).

Se han construido en muchos países bases de datos estratégicas para la realización de estudios bibliométricos. Precisamente la canadiense se ha empleado recientemente para realizar el análisis bibliométrico de la investigación sobre la obesidad financiada con cargo a los *Canadian Institutes of Health Research*. Los resultados de estos y otros trabajos son robustos, fiables y reproducibles. Se pueden aplicar a la caracterización de las actividades de investigación en un área y un contexto nacional y a su comparación con otros. De hecho, si el grupo de **Ramos** hubiera recuperado 2.713 trabajos en lugar de 1.407, la aportación española a la investigación de la obesidad en el mundo representaría una proporción diferente al 2,8%.

Cada vez es más frecuente que se publiquen protocolos de obligado cumplimiento para la realización de algunos tipos de estudios científicos, especialmente en el campo de la biomedicina. Los grupos españoles dedicados a la elaboración de indicadores y a la aplicación de la bibliometría al análisis de la actividad científica deberían acordar un protocolo y transmitirlo a los responsables de las revistas científicas españolas. Y no sólo a ellos: no hace ni cinco años se aprobaban en las universidades españolas tesis doctorales que analizaban la producción científica de todo un país cuantificando los trabajos extraídos de *Medline*.

Es bastante duro reconocer que la expresión “bibliometría de mesa camilla” (*poor man bibliometrics*) en la versión en español puede resultar desafortunada e hiriente, se pueda aplicar a buena parte de los estudios que se publican.

## Referencias bibliográficas

Bibliographic Services Division. *Authorship in Medline*. National Library of Medicine, 2010.

<http://www.nlm.nih.gov/pubs/factsheets/authorship.html>

Canadian Institutes of Health Research. *Bibliometric study of obesity research in Canada, 1998-2007* – CIHR. <http://www.cihr-irsc.gc.ca/e/41601.html>

**Belinchón-Carmona, Mercedes; Boada-Muñoz, Leticia; García-De-Andrés, Esther; Fuentes-Biggi, Joaquín; Posada-De-la-Paz, Manuel.** “Evolución de los estudios sobre autismo en España: publicaciones y redes de co-autoría entre 1974 y 2007”. *Psicothema*, 2010, v. 22, n. 2, pp. 242-249. <http://redalyc.uaemex.mx/pdf/727/72712496011.pdf>

**Casado-Mejía, Rosa; Ruiz-Arias, Esperanza; Solano-Parés, Ana.** “Características de la producción científica sobre cuidados familiares prestados por mujeres inmigrantes”. *Gaceta sanitaria*, 2009, v. 23, n.4, pp. 335-341.

[http://scielo.isciii.es/scielo.php?pid=S0213-91112009000400016&script=sci\\_arttext](http://scielo.isciii.es/scielo.php?pid=S0213-91112009000400016&script=sci_arttext)

**Costas, Rodrigo; Moreno, Luz; Bordons, María.** “Solapamiento y singularidad de *Medline*, *WoS* e *IME* para el análisis de la actividad científica de una región en Ciencias de la Salud”. *Revista española de documentación científica*, 2008, v. 31, n. 3, pp. 327-343. <http://redc.revistas.csic.es/index.php/redc/article/view/432/444>

**Falagas, Matthew E.; Pitsouni, Eleni I.; Malietzis, George A.; Pappas, Georgios.** “Comparison of *PubMed*, *Scopus*, *Web of Science*, and *Google Scholar*: strengths and weaknesses”. *FASEB journal*, 2008, v. 22, n. 2, pp. 338-342. <http://www.fasebj.org/content/early/2007/09/20/fj.07-9492LSF.full.pdf>

**Glänzel Wolfgang.** “The need for standards in bibliometric research and technology”. *Scientometrics*, 1996, v. 35, pp. 167-176. <http://www.springerlink.com/content/v11373p523h0m4v2/>

**Jiménez-Contreras, Evaristo; Torres-Salinas, Daniel; Ruiz-Pérez, Rafael; Delgado-López-Cózar, Emilio.** “Investigación de excelencia en España: ¿protagonistas o papeles secundarios?”. *Medicina clínica*, 2010, v. 134, n. 2, pp. 76-81. <http://eprints.rclis.org/handle/10760/13899>

**Lascurain-Sánchez, María Luisa; García-Zoritz, Carlos; Martín-Moreno, Carmen; Suárez-Balseiro, Carlos; Sanz-Casado, Elías.** “Impact of health science research on the Spanish health system, based on bibliometric and healthcare indicators”. *Scientometrics*, 2008, v. 77, n. 1, pp. 131-146. <http://www.akademai.com/content/p5822p2w054v4672/fulltext.pdf>

**Pestaña, A.** “El *Medline* como fuente de información bibliométrica de la producción española en biomedicina y ciencias médicas. Comparación con el *Science citation index*”. *Medicina clínica*, 1997, v. 109, n. 13, pp. 506-511.

**Ramos, José M.; Gutiérrez, Félix; Padilla, Sergio; Masiá, Mar; Martín-Hidalgo, Alberto.** “Geography of medical publications. An overview of HIV/AIDS research in 2003”. *AIDS*, 2005, v. 19, n. 2, pp. 219-220.

**Strotmann, Andreas; Zhao, Dangzhi.** “Combining commercial citation indexes and open-access bibliographic databases to delimit highly interdisciplinary research fields for citation analysis”. *Journal of informetrics*, 2010, v. 4, n. 2, pp. 194-200.

**Vioque, Jesús; Ramos, José-Manuel; Navarrete-Muñoz, Eva María; García-De-la-Hera, Manuela.** “Producción científica española en obesidad a través de *PubMed* (1988-2007)”. *Gaceta sanitaria*, 2010a, v. 24, n. 3, pp. 225-232. [http://scielo.isciii.es/scielo.php?pid=S0213-91112010000300009&script=sci\\_arttext](http://scielo.isciii.es/scielo.php?pid=S0213-91112010000300009&script=sci_arttext)

**Vioque, Jesús; Ramos, José-Manuel; Navarrete-Muñoz, Eva-María; García-De-la-Hera, Manuela.** “A bibliometric study of scientific literature on obesity research in *PubMed* (1988-2007)”. *Obesity reviews*, 2010b, v. 11, n. 8, pp. 603-11.

Register for free at <https://www.scipedia.com> to download the version without the watermark

<http://onlinelibrary.wiley.com/doi/10.1111/j.1467-789X.2009.00647.x/pdf>

\* \* \*

## Tratamiento de Scopus de los registros importados de Medline

Por Luis Rodríguez-Yunta

Me parece muy interesante plantear el debate de forma genérica sobre la fiabilidad de las fuentes utilizadas en bibliometría (no tanto discutir sobre un artículo concreto que opino que debería enviarse como carta a la revista que lo publicó).

Quería pedir a los autores una aclaración aprovechando que han investigado el tema. Conocía que Medline sólo refleja el lugar de trabajo del primer autor, pero tenía dudas sobre qué consecuencias tiene esto para los datos capturados por Scopus. Si Medline es utilizada para enriquecer la exhaustividad de Scopus, ¿Elsevier no se limita a descargar sus registros sino que enriquece los datos añadiendo las afiliaciones de los demás autores? ¿O lo hace sólo a partir de la ficha pregrabada de los autores (identificador de autor) que ya tuvieran otros registros en Scopus (es decir, a veces sí, a veces no, por ser un procedimiento automatizado)?

¿Hay información pública sobre cómo trata Scopus los registros importados de Medline?

[luis.ryunta@cchs.csic.es](mailto:luis.ryunta@cchs.csic.es)

de datos y se puede hacer un mal trabajo con esas mismas bases de datos. Hay que ser consciente de las limitaciones de cada una de ellas, evitarlas cuando no son las pertinentes para determinado análisis o complementarlas con otras bases de datos para obtener esa información por otra vía o crear una propia porque ninguna de las existentes ofrece la información necesaria para obtener determinados indicadores.

[aroldan@isciii.es](mailto:aroldan@isciii.es)

## La fiabilidad de Medline y su validez para estudios bibliométricos

Por José-Antonio Salvador-Oliván

Medline no es una fuente apropiada para realizar estudios bibliométricos que requieren el uso de determinadas variables no presentes en la base de datos, pero de ahí a decir que no es una fuente fiable... es mezclar churras con merinas. La fiabilidad tiene otro significado y se mide de diferente manera.

[jaso@unizar.es](mailto:jaso@unizar.es)

## Comentarios finales

Por Carlos B. Amat y Alfredo Yegros-Yegros

Muchas gracias por haber dedicado vuestro tiempo a contrastar vuestras opiniones con el contenido de nuestra nota y a enmendarla en sus errores. Sirva la siguiente lista de afirmaciones para mantener vuestro conocimiento y comentarios:

Desconocemos la mecánica por la cual muchos de los registros de Scopus, hasta los anteriores a 2004, tienen identificador de PubMed. Parece cierto que ha existido una cesión, por otra parte prevista, en las utilidades de PubMed.

<http://www.ncbi.nlm.nih.gov/books/NBK25501>

Suponemos que después de eso se ha realizado una conversión retrospectiva, quizá automática a partir de los registros originales de los editores. Por ejemplo, el documento con DOI 10.1097/GIM.0b013e31802d8353, que no corresponde a publicación de Elsevier alguna, aparece con sus afiliaciones intactas en el correspondiente registro de Scopus. No sucede así en todos los casos.

Así, en uno de los más notables (DOI 10.1126/science.1058040) tanto PubMed como Scopus pasan por alto la participación de dos investigadores del grupo de informática médica del Instituto Municipal de Investigación Médica (IMIM) de Barcelona y la Universidad Pompeu Fabra en los trabajos que condujeron a descifrar el genoma humano.

Valoramos Medline muchísimo y la empleamos continuamente como fuente en nuestros análisis

## Buenas y malas prácticas en el uso de bases de datos para análisis bibliométricos

Por Álvaro Roldán-López

Me parece un poco excesivo descalificar Medline como fuente para la realización de análisis bibliométricos con tanta rotundidad. Creo que dependerá del tipo de análisis bibliométrico que se quiera hacer.

Por poner un ejemplo, imaginemos que pertenezco a un Ciber y recabo de todos los grupos de investigación pertenecientes al mismo toda la producción de los últimos cuatro años, ¿podría hacer un análisis de la presencia de dicha producción en Medline? ¿Podría hacer un análisis de la temática de los grupos a través de los términos MeSH con los que han sido indizadas las publicaciones presentes en Medline?

Creo que el problema no es el uso de Medline sino el uso que se haga de esta base de datos. Es como intentar hacer un análisis temático partiendo de WoS. Creo que en general se puede hacer un buen trabajo bibliométrico con cualquier base

bibliométricos, aunque no como fuente única. En alguna parte de nuestro trabajo original figura la frase “cuando los estudios se realizan en un contexto nacional o comparan la producción científica de varios países, es imprescindible que las fuentes utilizadas expresen de forma completa la filiación institucional de los autores”.

Además de esa frase inicial en el resumen, nuestra nota aporta argumentos y algunas estimaciones que revelan las limitaciones de *Medline* en esos casos.

Por lo demás, hace ya tiempo que muchos grupos españoles que emplean la bibliometría como método de análisis (y a veces evaluación) de la actividad científica se han apuntado a la combinación de bases de datos de forma habitual. Las referencias que aportamos lo confirman y, si se quiere un ejemplo más reciente, echar un vistazo al prometedo original con DOI 10.1258/jhsrp.2010.009125.

Es más que posible que el título de la nota no exprese correctamente su contenido. Mejor sin duda “Los datos bibliométricos extraídos de registros de *Medline* no son fiables”<sup>1</sup>. En todo caso, la parte más importante de la nota es su último párrafo. A pesar de reconocer estas carencias, da que pensar el hecho de que casi no existan análisis basados exclusivamente en *Medline* en la literatura internacional.

Es cierto que no existe un premio Nobel de bibliometría, tampoco de matemáticas, ni de arquitectura. **Wolfgang Glänzel**, que ganó en 1999 la medalla *Derek John De Solla Price* de la *International Society for Scientometrics and*

*Informetrics*, es quien defiende la necesidad de compatibilidad entre fuentes para asegurar la comparabilidad de resultados. Una cuestión meramente técnica, la califica él en el trabajo al que hacemos referencia.

No sabemos si es una lástima que *Medline* no sirva para la cibernetría, de hecho no sabemos si sirve o no. Tampoco sabemos qué opinan los del *open access* sobre todo esto. Sí que estamos seguros de que las limitaciones de *Medline* en relación con la filiación institucional de los investigadores desaparecerán, del mismo modo que han ido desapareciendo las limitaciones en el número de autores que se incluían en cada registro.

<http://www.nlm.nih.gov/bsd/mms/medlineelements.html#au>

Reiteramos nuestro agradecimiento a quienes han comentado la nota, pero no dejamos de reconocer el fracaso de su publicación. El párrafo final quería expresar un llamamiento primero al acuerdo y luego a la plasmación de ese acuerdo en un documento de consenso que, claro, ya no será posible elaborar. Es posible que un foro distinto a *ThinkEPI* resultara más indicado, pero el llamamiento que dirigimos para aclarar lo que se ha dado en llamar “citable items” a la lista *Incyt* sobre indicadores de ciencia y tecnología solo obtuvo una respuesta procedente de Bogotá.

#### Nota

1. El título original de la nota era “*Medline* no es una fuente fiable de análisis bibliométricos”, modificado

Register for free at <https://www.scipedia.com> to download the version without the watermark